

Feature Learning from Massive Spatial Trajectories: A Case Study of Map Matching

Jian Yang,¹ Linfang Ding,^{2,3} Liqiu Meng,³ and Xiong You¹

1. Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China; jian.yang@tum.de, youarexiong@163.com

2. Universitaet Augsburg, Augsburg, Germany; linfang.ding@geo.uni-augsburg.de

3. Technische Universitaet Muenchen, Munich, Germany; liqiu.meng@bv.tum.de

Keywords: Feature Learning, Trajectory Data Mining, Map Matching, Conditional Random Fields

Mining spatial trajectories aims to extract non-explicit information from spatial trajectory data that can be organized as temporally ordered locations, such as taxi GPS logs, twitter check-ins (Zheng, 2015). The field has been revolutionizing the traditional means of collecting and processing geo-spatial information for mapping and many other real-world applications (Bock, Liu, & Sester, 2016; Yang & Meng, 2015). One of the mining tasks requires the labeling of individual points in trajectories with states in query such that the physical measurements can be better interpreted (Yang, 2016). For example, by means of map matching, each data point in the location sequence is assigned to the road segment on which the moving object traveled, while methods of location-based activity recognition are used to identify the most probable activities (e.g., at home, at work, at bar) associated with each location in the trajectory data. These labeling tasks impose challenges on label assignments especially when the measurements are noisy and when there are non-exclusive semantic correspondences between data points and labels.

Probabilistic methods are popular in solving these labeling tasks as they often produce better label accuracies. In a previous work (Yang, 2016), the labeling of spatial trajectories based on map matching was treated. We developed a probabilistic model with conditional random fields, which computes the maximum likelihood of the trajectory data given label assignments based on a set of weighted features that captures the correspondences between road segments and location observations of the moving objects. On a small taxi GPS dataset, our model outperformed the state-of-the-art approaches in terms of both accuracy and reliability of matching GPS taxi trajectories at a low sampling rate to OpenStreetMap road data. Furthermore, our model employed an optimization process to select the most relevant features (i.e., features that improve the model likelihood for map matching, see Fig.1) for matching sparse and noisy GPS trajectories, some of which revealed valuable cues to understand drivers' driving behavior at road intersections.

However, some interesting questions remain unanswered in our current results: 1) How to interpret the features learned from sample taxi routes? 2) Would the overall confidence (i.e., the likelihood) of the model for the route prediction suggest the correct matching result? 3) How the routing preference of an individual taxi driver could deviate from the collective knowledge mined from massive trajectory data?

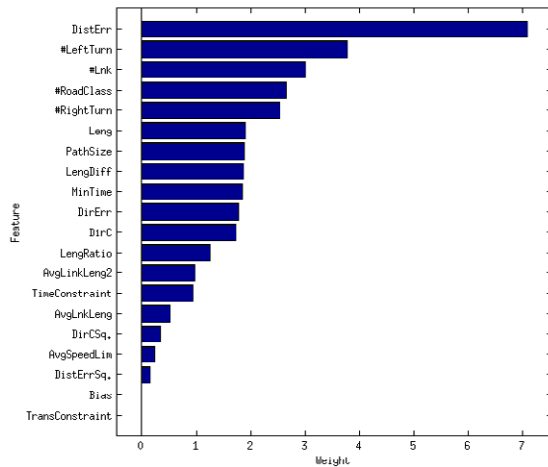


Fig. 1. Features learned for map matching of low sampling rate GPS data. [The weights' magnitudes indicate the relevance degree of the feature to the task. Among all the features, distance error (DistErr), number of left turn (#LeftTurn), number of the link in the path (#Lnk) and number of different road classes in the path (#RoadClass) are the most relevant ones. (Source: Yang, 2016)]

Jian Yang 2/28/2017 3:06 AM

Comment [1]: Rephrase a little bit

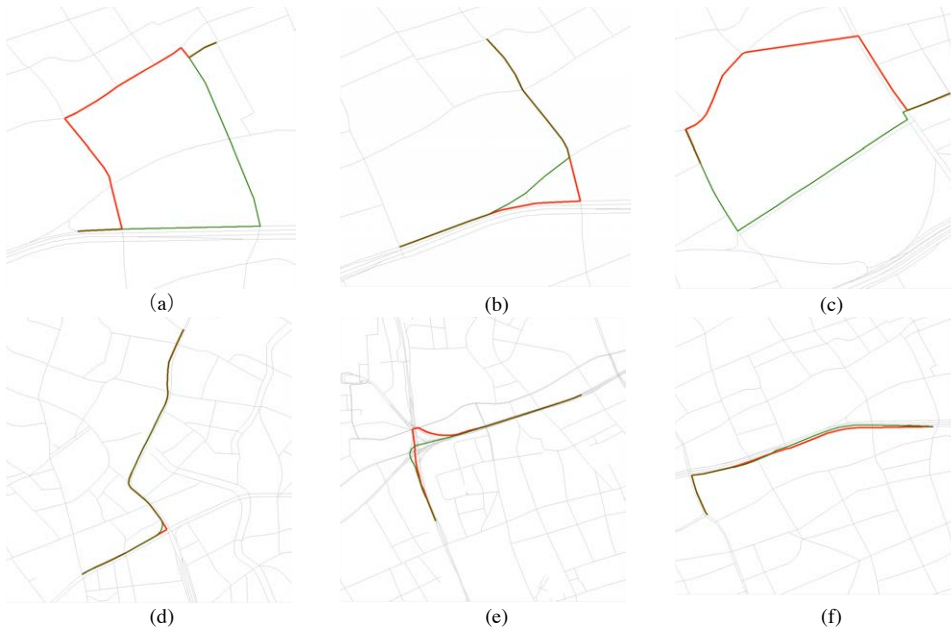


Fig. 2. Recovered paths between GPS data points with sampling rate of 120s. Green paths are ground truths and red ones are results generated by our map matching method. The comparisons illustrate the cases when fastest paths are less preferred by the taxi drivers: (a) path with fewer turns, (b)-(c) path skipping traffic crossing, (d)-(e) paths with smooth transitions, (f) path with fewer lane transitions.

To investigate aforementioned research questions, this paper scales up the original map matching. Firstly, the ground truth data are prepared in a semi-automated manner. Since ground truth data are often not available in massive trajectory dataset, a trajectory data management system is developed to match trajectory data with a high sampling rate using Hidden Markov Model (Newson & Krumm, 2009), followed by a carefully designed manual visual validation to ensure the quality of the ground truth data. Secondly, visual analytics approaches (Ding, 2016) are proposed to explore spatio-temporal patterns of individual routing preferences (Fig.2 shows that drivers are not taking shortest path, fastest path in some cases), namely when individual drivers deviate from the fastest route and how often they make these decisions. Thirdly, we train our chain structured conditional random fields with these labeled data. And experiments, e.g. with the examination of the turn-by-turn patterns at different road intersections, are performed to interpret features against daily driving experiences.

Based on this extensive study, a number of conclusions incl. our new insight can be drawn: 1) The quality of labeled data has a significant impact on the feature learning results and the performance of the probabilistic models; 2) The interpretation of learned features should be carefully used to understand routing preference.

References

- Bock, F., Liu, J., & Sester, M. (2016). Learning On-Street Parking Maps from Position Information of Parked Vehicles. In *Geospatial Data in a Changing World* (pp. 297–314). http://doi.org/10.1007/978-3-319-33783-8_17
- Ding, L. (2016). *Visual Analysis of Large Floating Car Data - A Bridge-Maker between Thematic Mapping and Scientific Visualization*.
- Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, 336.
- Yang, J. (2016). *Labeling Spatial Trajectories in Road Network Using Probabilistic Graphical Models*. Technischen Universität München.
- Yang, J., & Meng, L. (2015). Feature Selection in Conditional Random Fields for Map Matching of GPS Trajectories. In G. Gartner & H. Huang (Eds.), *Progress in Location-Based Services 2014, Lecture Notes in Geoinformation and Cartography* (pp. 121–135). Springer International Publishing. http://doi.org/10.1007/978-3-319-11879-6_9
- Zheng, Y. (2015). Trajectory Data Mining: An Overview. *ACM Transaction on Intelligent Systems and Technology*, 6(3). Retrieved from <http://research.microsoft.com/pubs/241453/TrajectoryDataMining-tist.pdf>